BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 6: e94039
doi: 10.3897/biss.6.94039

OPEN ACCESS

Conference Abstract

Building a Database using Unconventional Sources: Squirrels of India

Swati Udayraj‡, Senan D'Souza‡, Aravind P S‡, Nandini Rajamani‡

‡ IISER Tirupati, Tirupati, India

Corresponding author: Nandini Rajamani (nandini@labs.iisertirupati.ac.in)

Received: 25 Aug 2022 | Published: 07 Sep 2022

Citation: Udayraj S, D'Souza S, P S A, Rajamani N (2022) Building a Database using Unconventional Sources: Squirrels of India. Biodiversity Information Science and Standards 6: e94039.
https://doi.org/10.3897/biss.6.94039

Abstract

Squirrels, like most other small mammals, have been poorly documented in the Indian subcontinent, which deters us from understanding species declines and prioritizing research on sensitive taxa (McKinney 1999, Koprowski and Nandini 2008). They are diverse in their habits and morphology and perform important ecological roles, including seed dispersal, pollination, and regulating plant growth. They also form a significant prey base for many predators. Squirrels respond strongly to pressures around them, including urbanization, habitat modification, and climate change, and can thus act as model study systems (Sol et al. 2013). The first step in understanding how species respond to such changes is by gaining knowledge of where they occur.

The rapid spread of internet connectivity and access to mobile technology across India allow us to access large-scale secondary data in ways that were not possible around a decade ago. We created a pan-India database for 30 species of squirrels using primary data (from fieldwork) and secondary sources (museum records, published and gray literature), citizen science portals (six sources), and social media platforms (14 sources). The use of social media platforms is increasing exponentially across India, yet these remain a largely unexplored source for harvesting biodiversity information. Given low public awareness of squirrel species, we expected high error rates with contributors' assignment of species identity. A key of species images and calls was used while data gathering to maintain consistency across the team. A pipeline for the data collection and curation was

© Udayraj S et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

created, and all volunteers on the project were trained to maintain consistency in data collection. To ensure verification of species identity, media (photographs, audio, and video data) are collected when possible or are cross-checked on the source site. Some (iNaturalist, Project Noah) citizen science platforms allow script-based or search-based downloads of bulk records without media. Each media record on such citizen science sites is manually checked to confirm species identification.

On social media platforms (Fig. 1), species-wise searches were performed (using common and scientific names) within each platform. For all social media records, media (photograph, audio, video) data was downloaded along with location, date, observer, and relevant notes. Each entry was manually entered into a database by researchers (12 over two years, including volunteer interns). Each record was then manually verified for species by one or two of four curators (more curators for less-familiar species). Duplicates were manually removed by a curator, who compares species-specific data across multiple sources. The location for each entry was also curated, and a georeference was added when unavailable in the original post. All location data were imported into Google Sheets, and the map tool Geocode by Awesome Table were used to obtain latitude and longitude data for places. On many occasions, curators contacted observers on social media to confirm details before an entry was finalized.

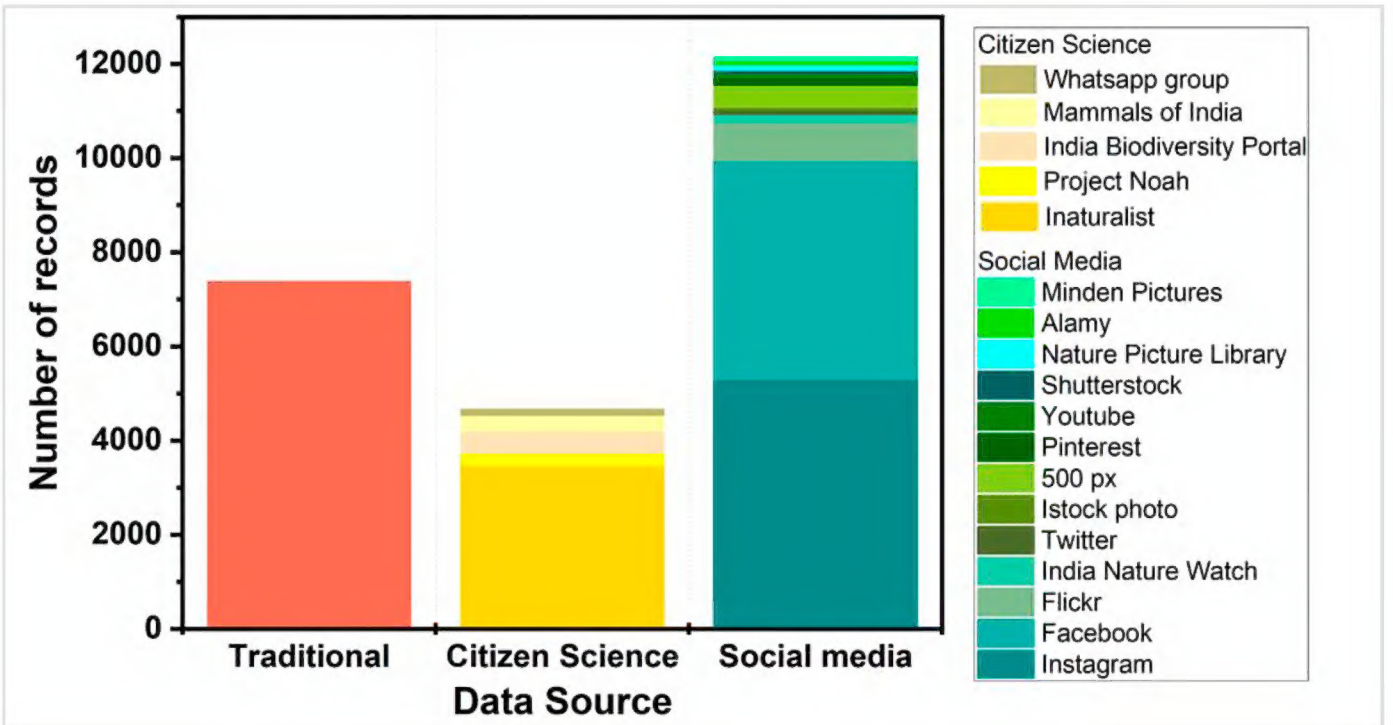

Figure 1.

Source-wise occurrence records collected for the Squirrels of India database.

Over two years, the database grew to include 24,170 records with approximately 14,000 media files, with the team working for over 2200+ hours. About 48% (12,035) of the occurrence records came from social media sources, followed by 30% of records (7375) from traditional sources and 22% (4660) from citizen science portals (Fig. 1).

We examined the temporal trends and bias for squirrel occurrence data for all three sources and assess the over and under-representation of squirrel occurrence based on

body size, activity period, body-color, International Union for Conservation of Nature (IUCN) Red List status, range size, and habitat type.

The majority of the occurrence records were that of tree squirrels (Fig. 2), followed by flying and ground squirrels. This is likely because tree squirrels are diurnal and more abundant, and hence are easier to record when compared to flying squirrels which are cryptic and nocturnal. The two species of ground squirrels in India are restricted to higher elevations in the Himalayas, making them difficult to record.

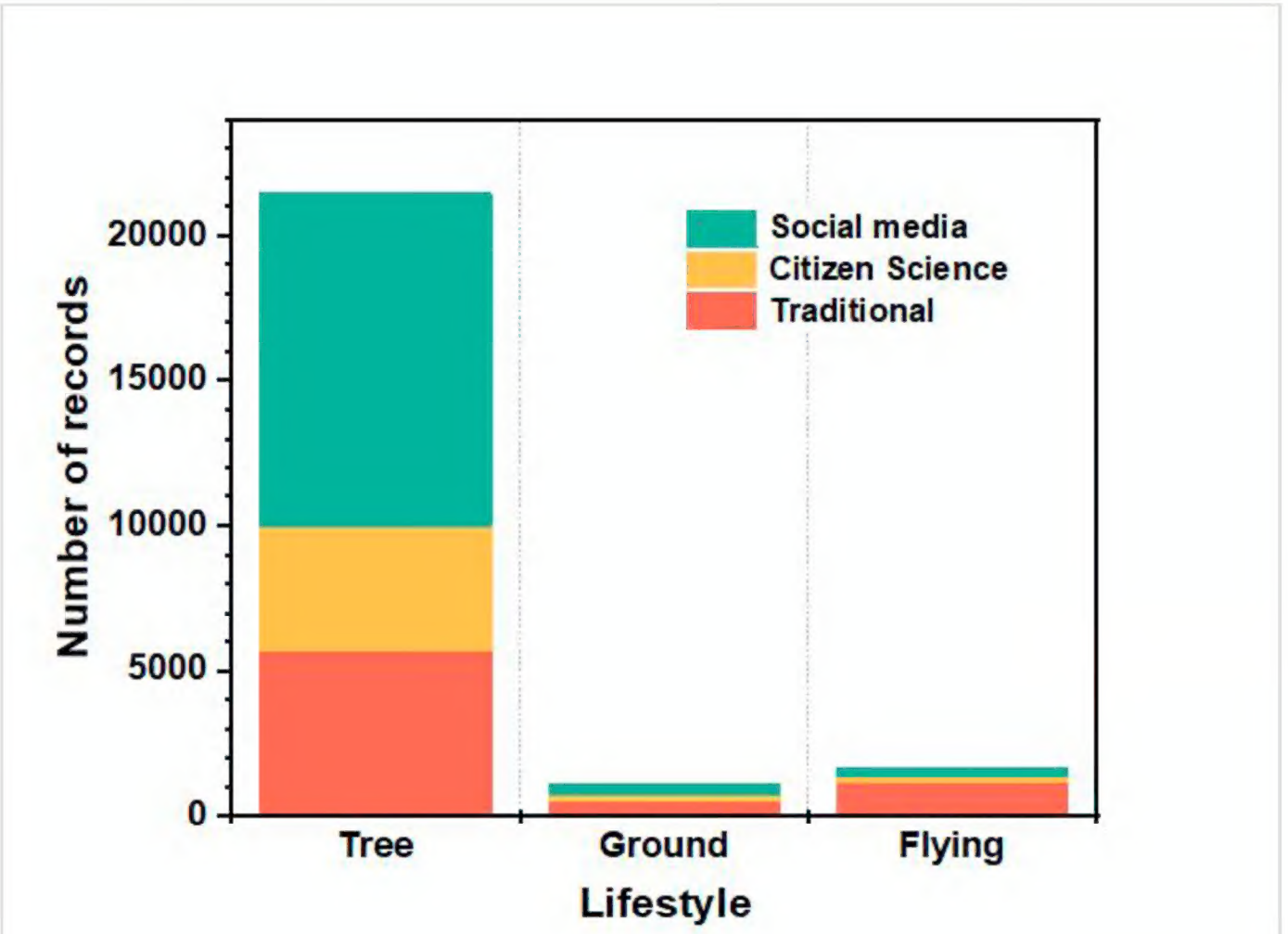

Figure 2.

Source-wise occurrence records for squirrels based on their lifestyle ($\chi 2$ =1695.11, df = 4, *p*-value < 0.00001).

There are, however, differences in records across regions in India. Based on a quick examination of the occurrence records, most of them are from urban areas, reflecting either bias in data collection (concentrated human densities) or species response to urbanization. Some species like *Funambulus palmarum* and *Funambulus pennantii* are known to be abundant in areas with higher human densities, which might be reflected in the number of occurrence records. In contrast, most other species seem restricted to areas with less anthropogenic disturbance. Therefore, recording fine-scale occurrences for this diverse group is crucial to understand species' responses to rapid landscape modifications such as urbanization.

Our understanding of biodiversity in a changing world has been greatly improved by combining, harmonizing, and analyzing large amounts of heterogeneous ecological data (Hampton et al. 2013). The availability of more accurate data enables studies to address questions at increasingly large spatial and temporal scales with stronger inference and more accurate and predictive models, which, in turn, yield important biological insights (Lewis et al. 2018).

Keywords

social media, occurrence, media

Presenting author

Swati Udayraj

Presented at

TDWG 2022

Hosting institution

IISER TIRUPATI

References

- Hampton SE, Strasser CA, Tewksbury JJ, Gram WK, Budden AE, Batcheller AL, Duke CS, Porter JH (2013) Big data and the future of ecology. Frontiers in Ecology and the Environment 11 (3): 156-162. https://doi.org/10.1890/120103
- Koprowski JL, Nandini R (2008) Global hotspots and knowledge gaps for tree and flying squirrels. Current Science 95 (7): 851-856. URL: https://www.jstor.org/stable/24103178
- Lewis K, Wal EV, Fifield D (2018) Wildlife biology, big data, and reproducible research. The Wildlife Society https://doi.org/10.1002/wsb.847
- McKinney M (1999) High rates of extinction and threat in poorly studied taxa. Conservation Biology 13 (6): 1273-1281. https://doi.org/10.1046/j.1523-1739.1999.97393.x
- Sol D, Lapiedra O, González-Lagos C (2013) Behavioural adjustments for a life in the city. Animal Behaviour 85 (5): 1101-1112. https://doi.org/10.1016/j.anbehav.2013.01.023